FREESTYLE MARKUP LANGUAGE

Denis Pondorf
Universität Bremen, Bremen, Germany

Andreas Witt
Institut für Deutsche Sprache, Mannheim, Germany

Aug. 5, 2010

Balisage Conference, Montréal, Canada

table of content

- **introduction**
- **architecture**
- **concepts**
- **status**
- **future work**

INTRODUCTION: What is FML ?

- markup (meta-) language

  (extensible, generalized, descriptive, open)

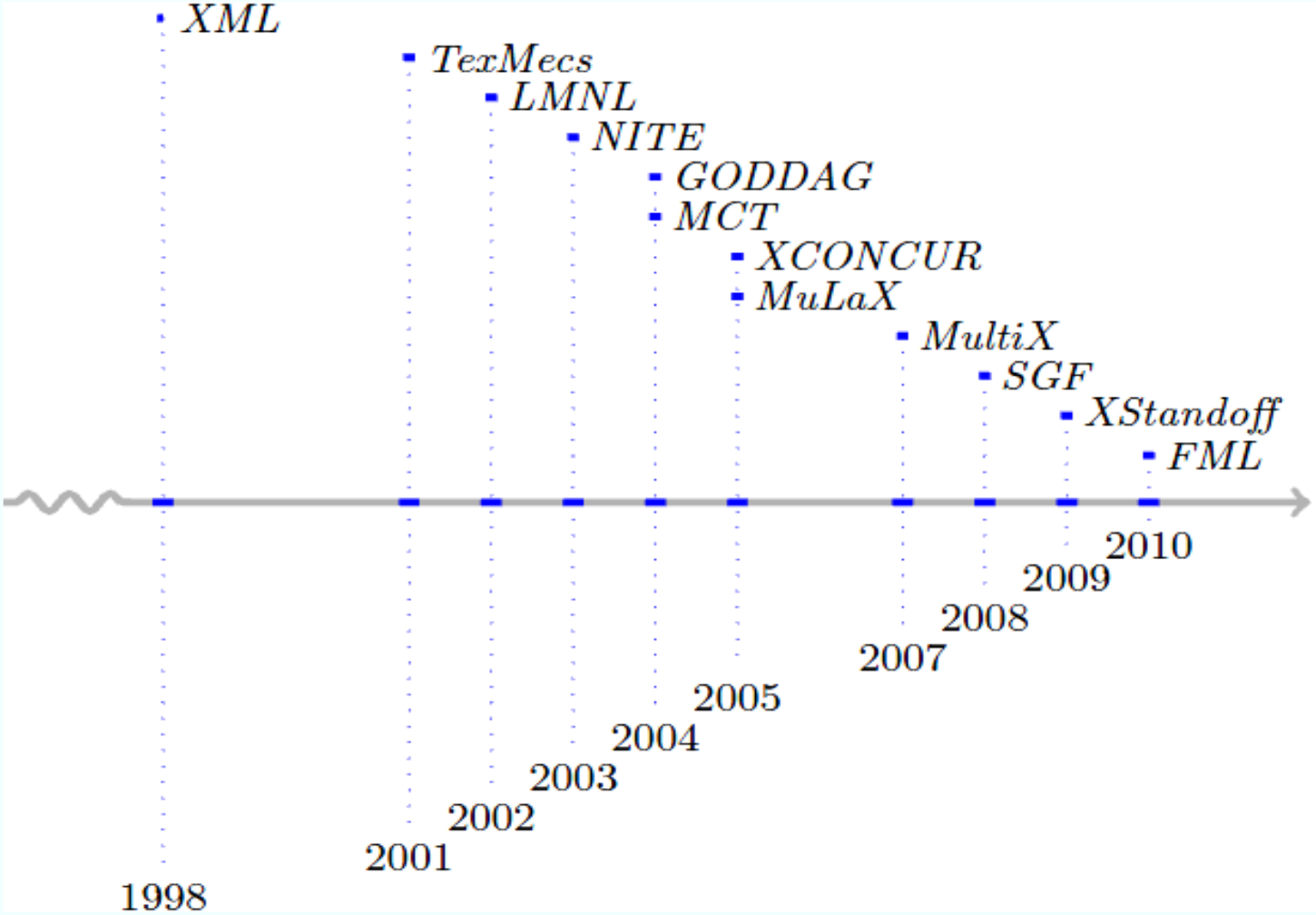
- simplified and extended XML

- principles: „maximum freedom“
  „markup data container“
  „simplicity“
  „document ⌦ transformation ⌦ graph“

INTRODUCTION: *hello world* example

```
<@fml.name="introduction“
  fml.uri="http://www.freestyle-markup.org/intro.fml"
  fml.description="fml example document">

supercalifragilisticexpialidocious

<! translation: "wonderful" !>

<movie type="musical">
      Mary Poppins
</movie>

<year> 1964 </year>
```

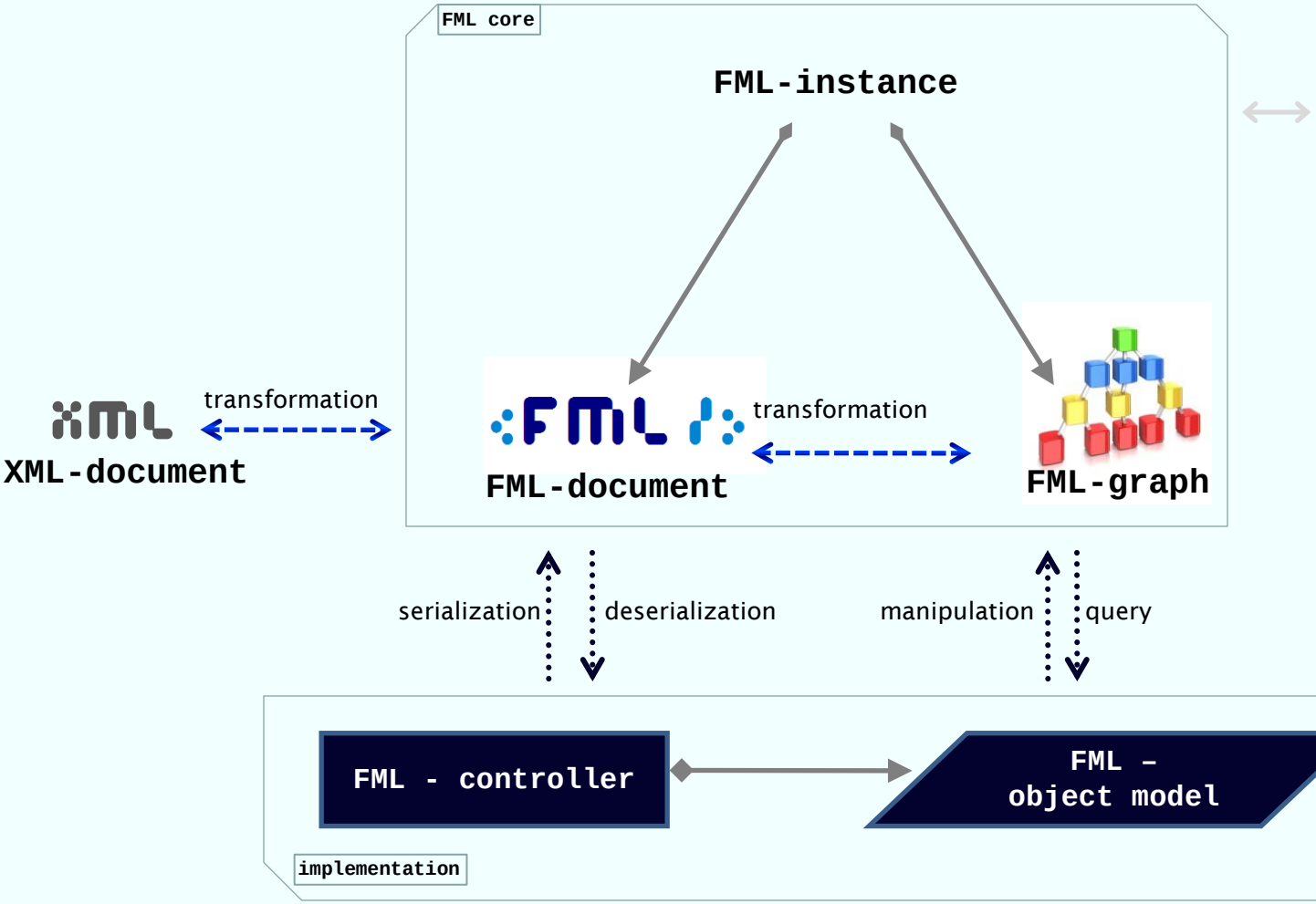
INTRODUCTION: related work



INTRODUCTION: Why FML ?

- deficit discourse consolidation:
  - consolitates existing approaches
  - introduces new approaches
- 10 concepts in one solution
- 19 requirements ➔ 19 features
- relevant integration scenarios (CEBIT-survey)
- ready2use
- pushing evolution

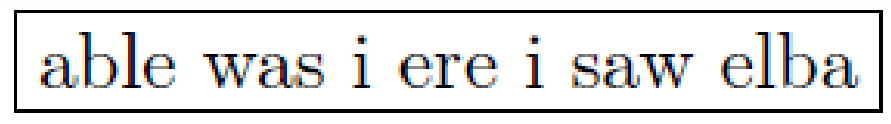
ARCHITECTURE



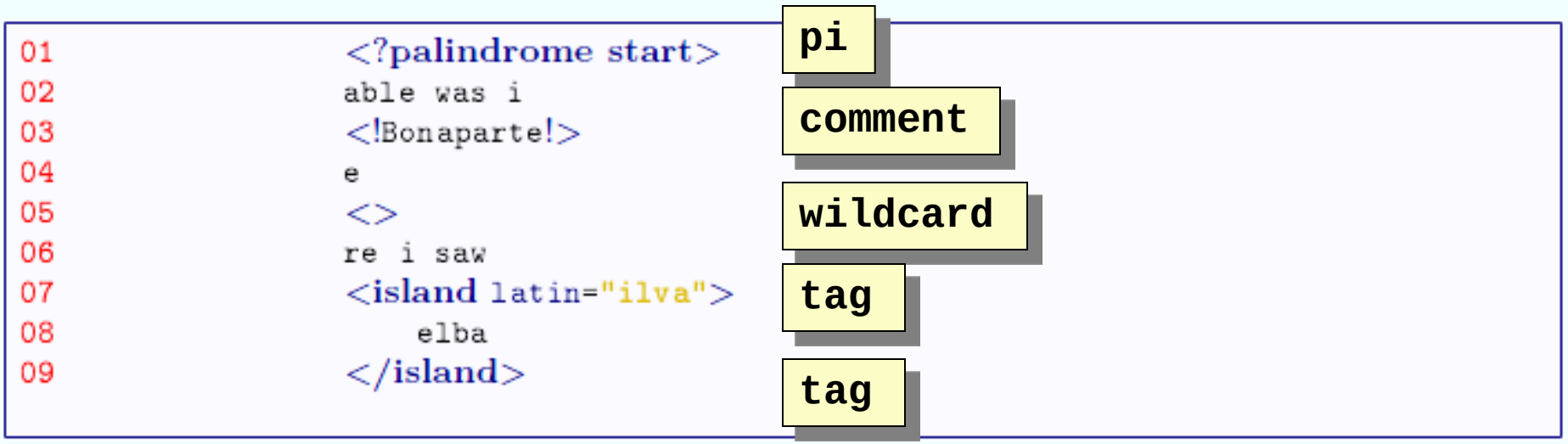
CONCEPTS

- **annotation**
- **declaration**
- **tagging**
- **attribution**
- **interference**
- **identification**
- **congruence**
- **independence**
- **segmentation**
- **fragmentation**

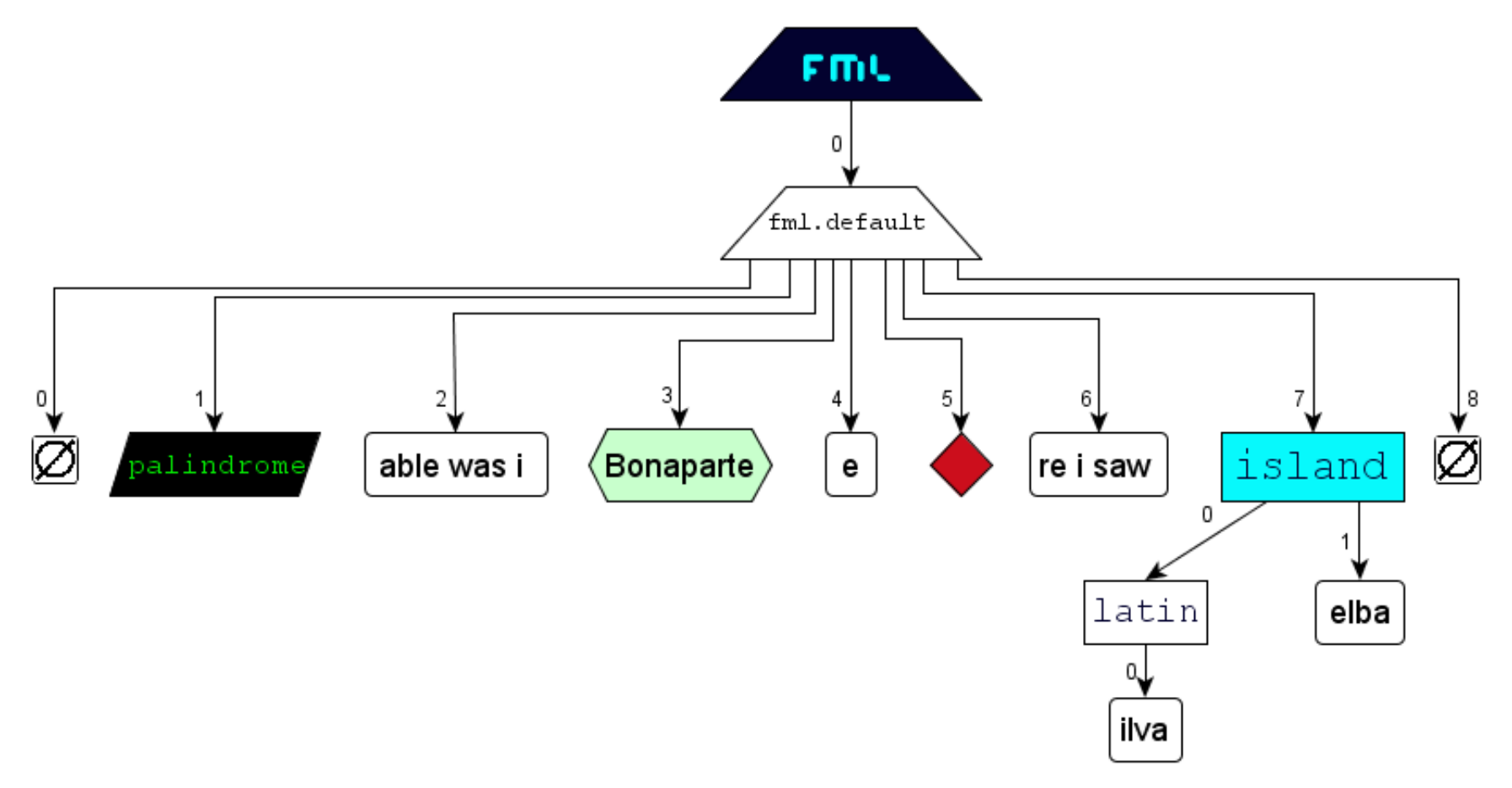
CONCEPTS: annotation



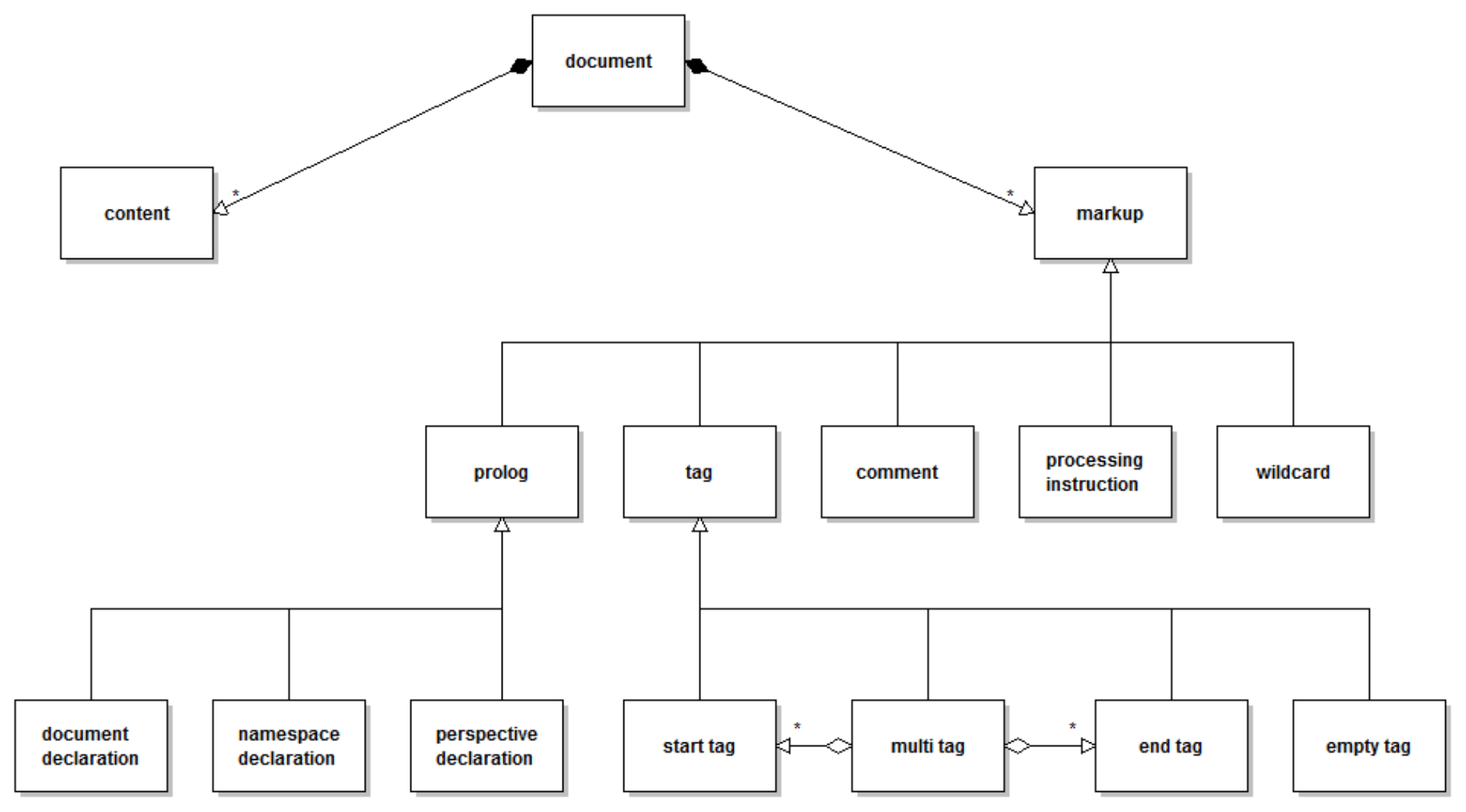
CONCEPTS: annotation



CONCEPTS: annotation



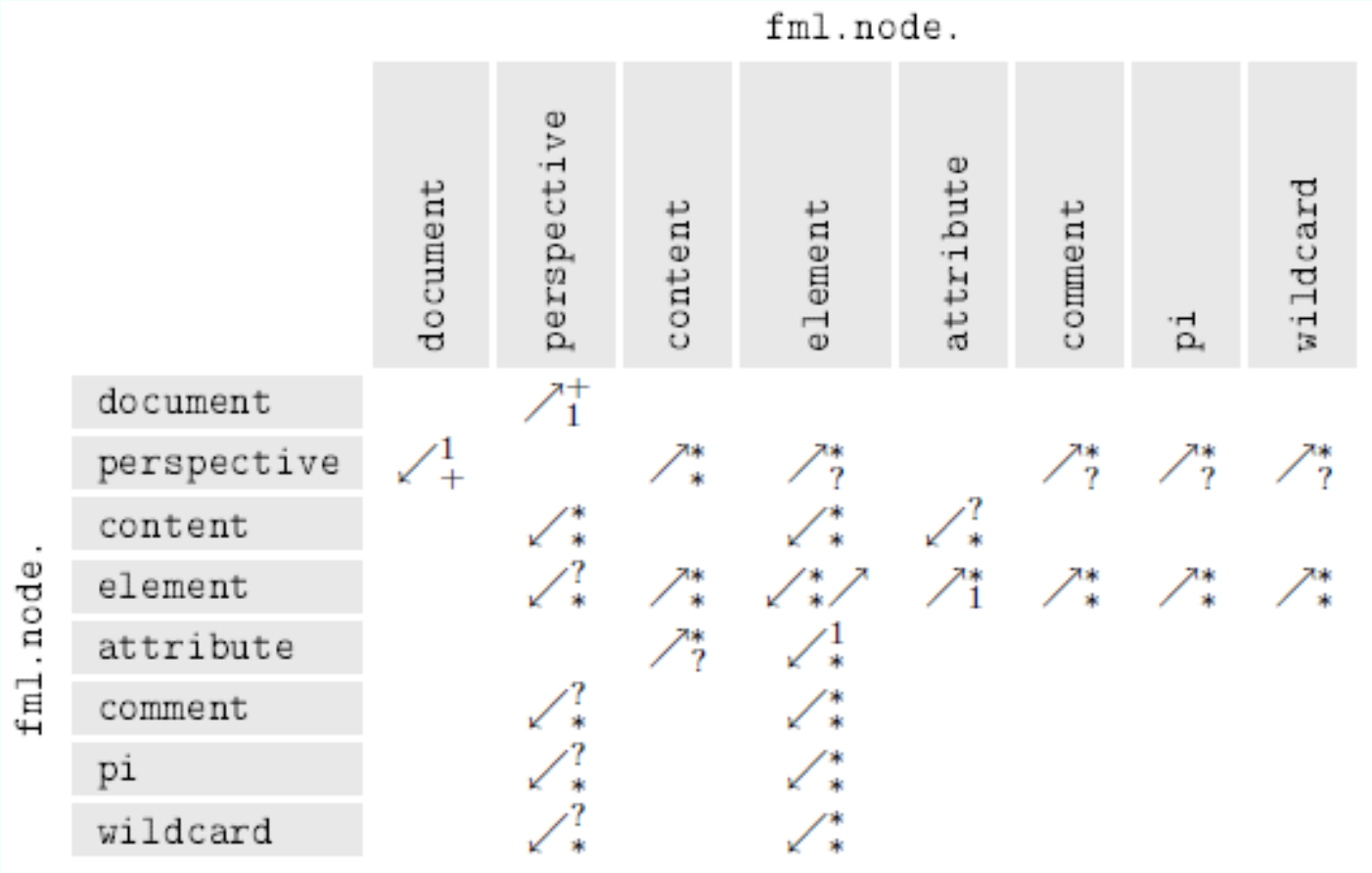
COMPONENTS


document
content
markup
prolog
tag
comment
processing instruction
wildcard
document declaration
namespace declaration
perspective declaration
start tag
multi tag
end tag
empty tag
*
*
*
*

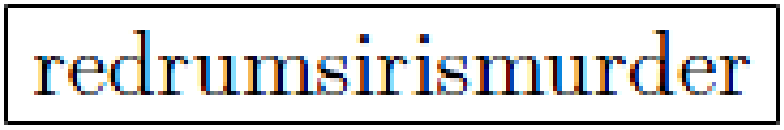
EBNF - GRAMMAR

```
fml.document = fml.prolog? (fml.content | fml.tag | fml.comment | fml.pi | fml.wildcard)* ;
fml.prolog = fml.prolog.document fml.prolog.perspective* fml.prolog.namespace* ;
fml.prolog.document = '<@' ('fml.name="' fml.attribute.value '"') (space 'fml.uri="' fml.attribute.value '"')? (space 'fml.descr
(space 'fml.fragment="' fml.attribute.value '"')? (space 'fml.schema="' fml.attribute.value '"')? (space 'fml.trim="' ('true'|'f
fml.prolog.perspective = '<@' ('fml.perspective.name="' fml.attribute.value '"') (space 'fml.perspective.uri="' fml.attribute.va
'fml.perspective.schema="' fml.attribute.value '"')? '>' linewrap ;
fml.prolog.namespace = '<@' ('fml.namespace.name="' fml.attribute.value '"') (space 'fml.namespace.uri="' fml.attribute.value '"
fml.content = ( (UTF-8-character - '<') | '\<' )* ;
fml.tag = '<' (fml.tag.start | fml.tag.end | fml.tag.empty | fml.tag.multiple) '>' ;
fml.tag.start = (fml.perspective.name '|')? (fml.namespace.name ':')? fml.tag.name (fml.tag.id)? (space fml.attribute)* ;
fml.tag.end = (fml.perspective.name '|')? '/' (fml.namespace.name ':')? fml.tag.name (fml.tag.id)? ;
fml.tag.empty = (fml.perspective.name '|')? (fml.namespace.name ':')? fml.tag.name (space fml.attribute)* '/' ;
fml.tag.multiple = (fml.tag.start | fml.tag.end | fml.tag.empty) (fml.tag.start | fml.tag.end | fml.tag.empty)+ ;
fml.tag.name = fml.name ;
fml.tag.id = '#' fml.name ;
fml.perspective.name = fml.name ;
fml.namespace.name = fml.name ;
fml.attribute = fml.attribute.name '="' fml.attribute.value '"' (',"' fml.attribute.value '"')* ;
fml.attribute.name = fml.name ;
fml.attribute.value = ( (UTF-8-character - '"') | '\"' )* ;
fml.comment = '<!' fml.comment.content '!>' ;
fml.comment.content =  | UTF-8-character | (UTF-8-character? ( ('!' (UTF-8-character - '>') ) | ( (UTF-8-character - '!') UTF-8-
fml.pi = '<?' (fml.perspective.name '|')? fml.pi.target space fml.pi.instruction '>' ;
fml.pi.target = fml.name ;
fml.pi.instruction = ( (UTF-8-character - '>') | '\>' )+ ;
fml.wildcard = '<' (fml.perspective.name '|')? '>' ;
fml.name = ( (UTF-8-character - escape-symbols.exclude) | escape-symbols.include )+ ;
UTF-8-character = [U+0000 - U+FFFF] ;
space = U+0020 ;
linewrap = U+000A ;
escape-symbols.exclude = ( '>', '<', '\', '@', '?', '!', '/', '|', ':', '#', space ) ;
escape-symbols.include = ( '\>' | '\<', '\\', | '\@' | '\?' | '\!' | '\/' | '\|' | '\:' | '\#' | '\' space ) ;
```

GRAPH

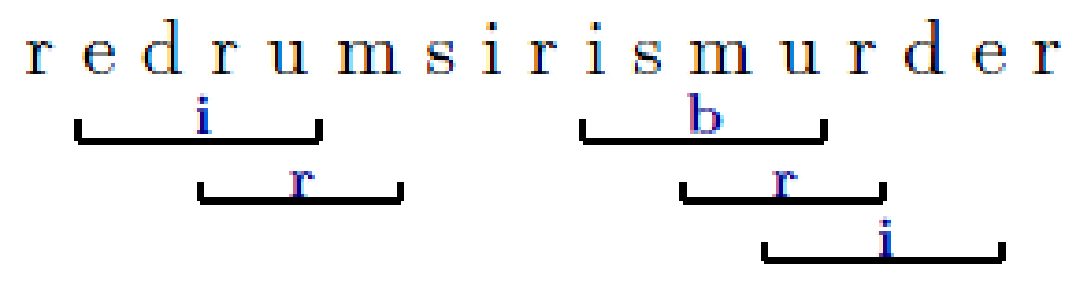

fml.node.
document
perspective
content
element
attribute
comment
pi
wildcard
fml.node.
document
perspective
content
element
attribute
comment
pi
wildcard

CONCEPTS: interference



CONCEPTS: interference

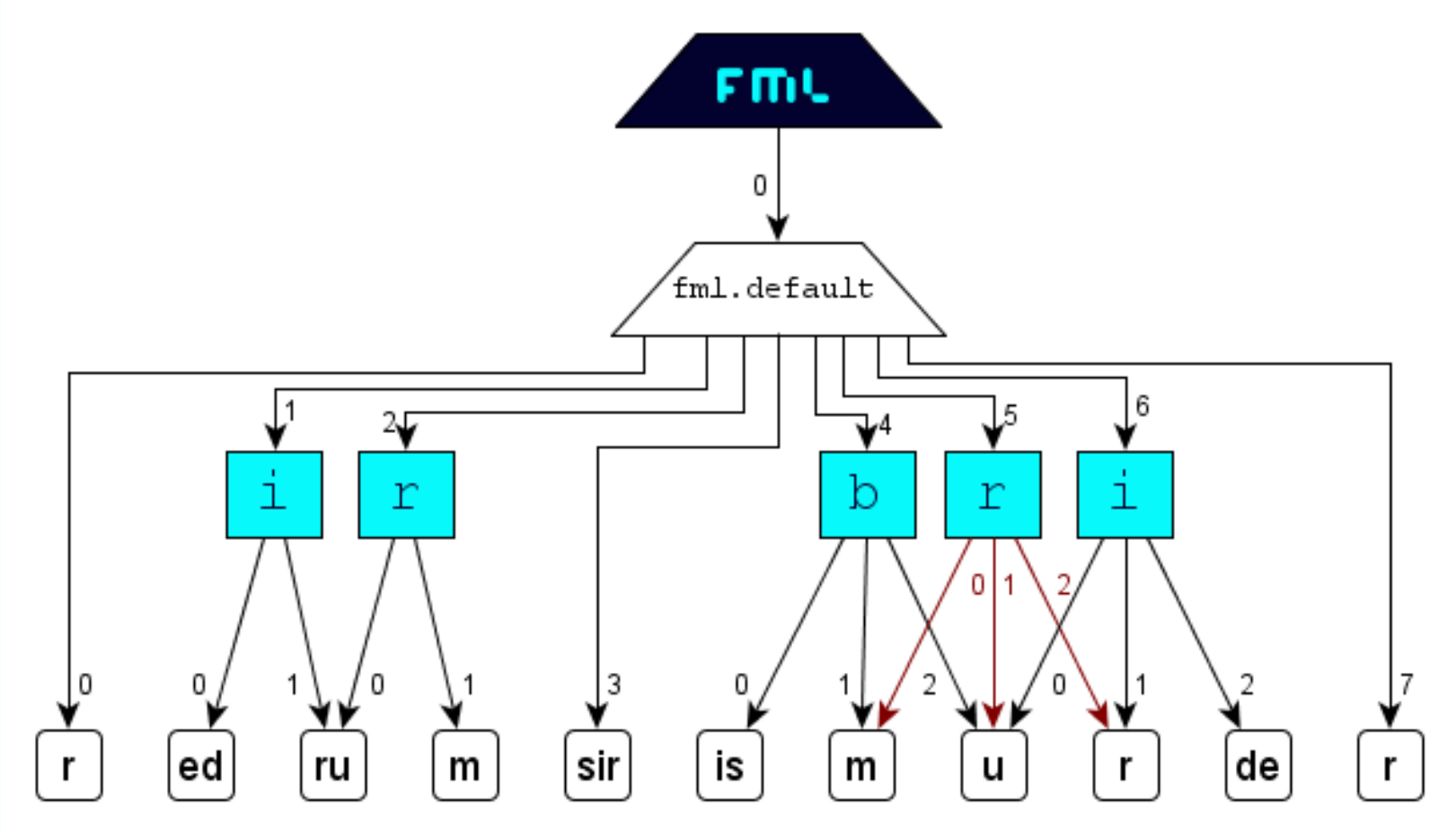
CONCEPTS: interference



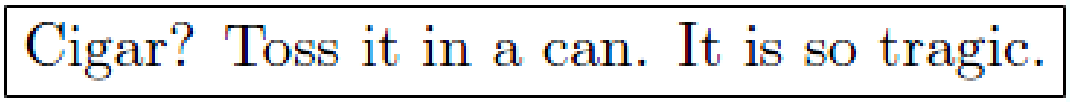
CONCEPTS: interference

r e d <r> r u m </r> s i r i s <r> m u r </r> d e r

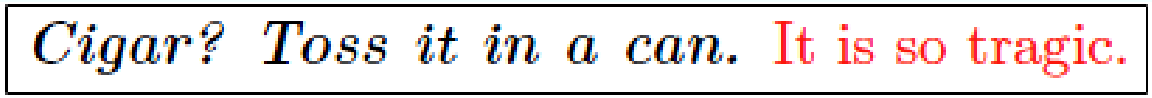
CONCEPTS: interference



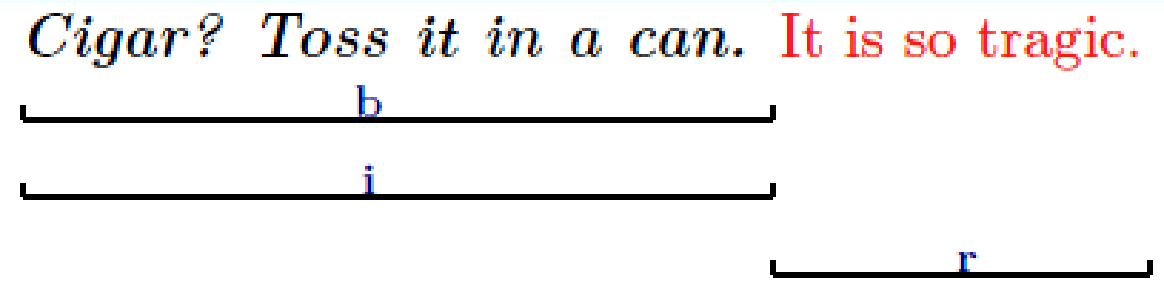
CONCEPTS: congruence



CONCEPTS: congruence



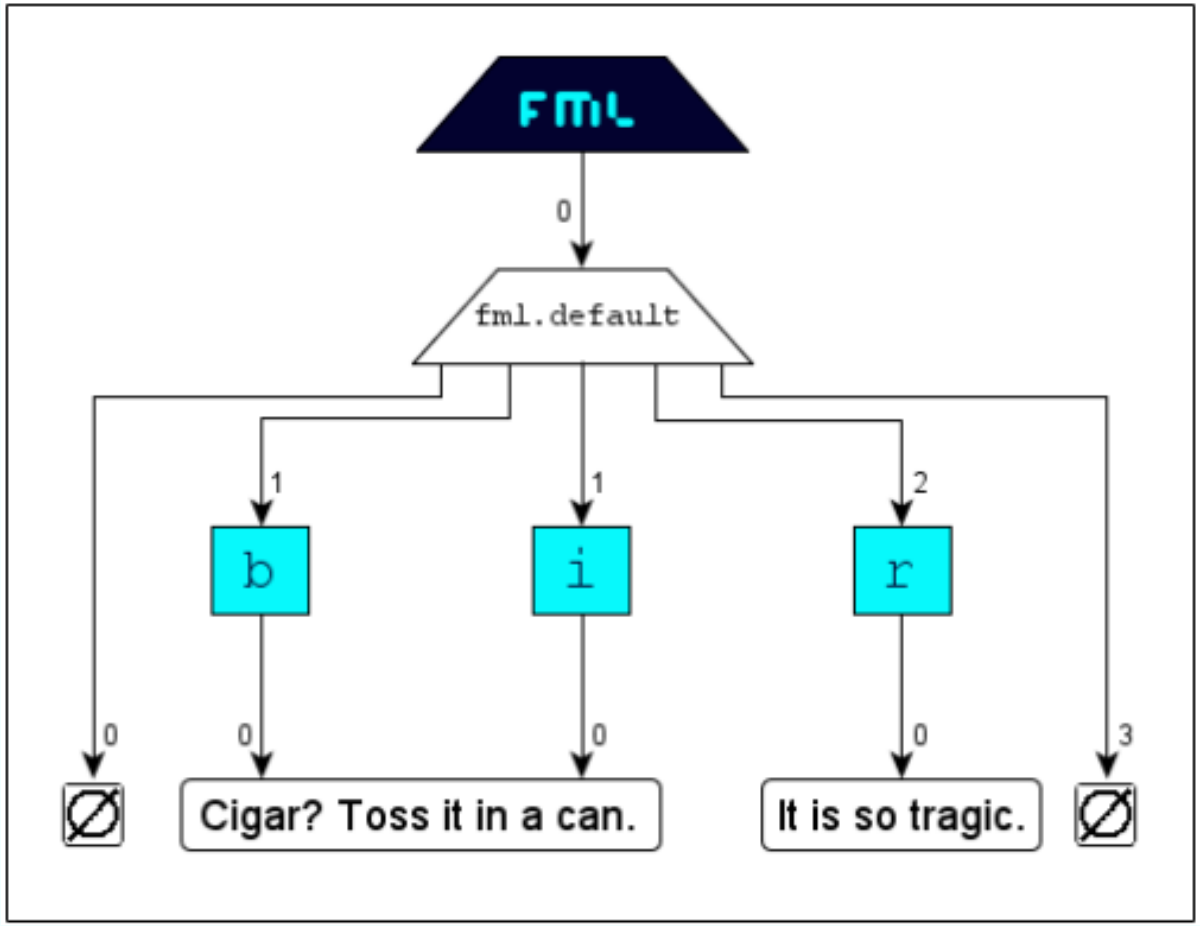
CONCEPTS: congruence



CONCEPTS: congruence

```
<b i>Cigar? Toss it in a can. </b /i  r>It is so tragic.</r>
```

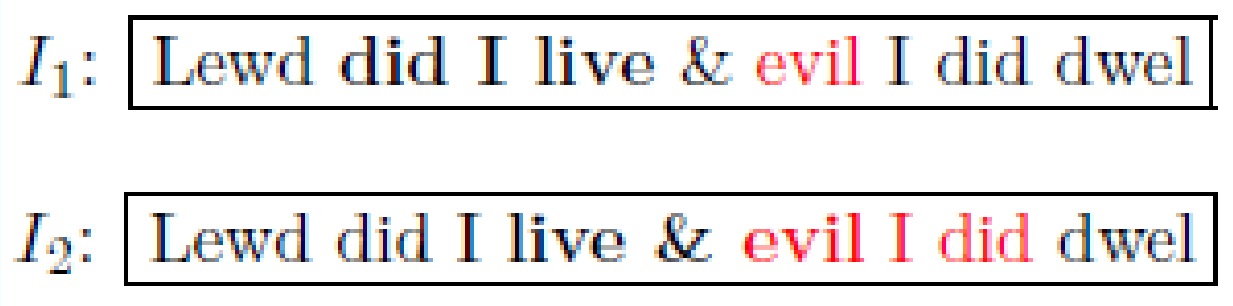
CONCEPTS: congruence



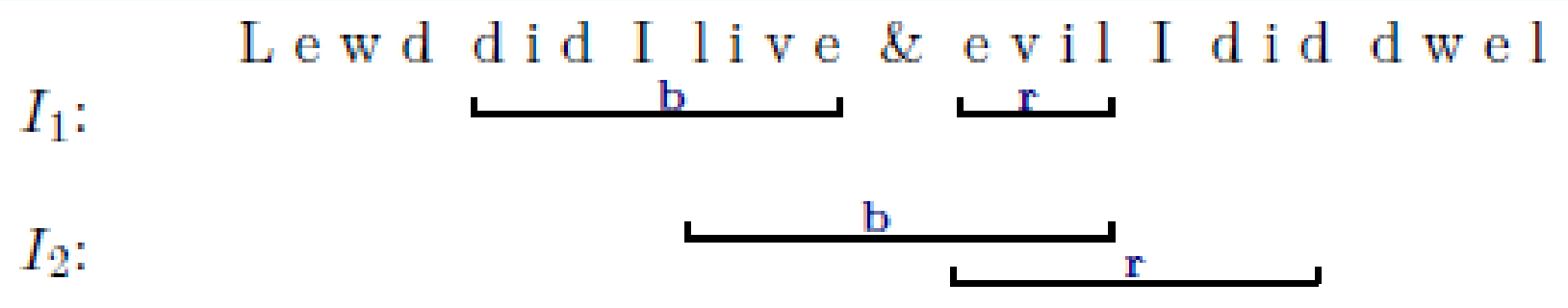
CONCEPTS: independence

(multiple annotations, multiple perspectives, views, layers, multi-rooted trees, ...)

CONCEPTS: independence



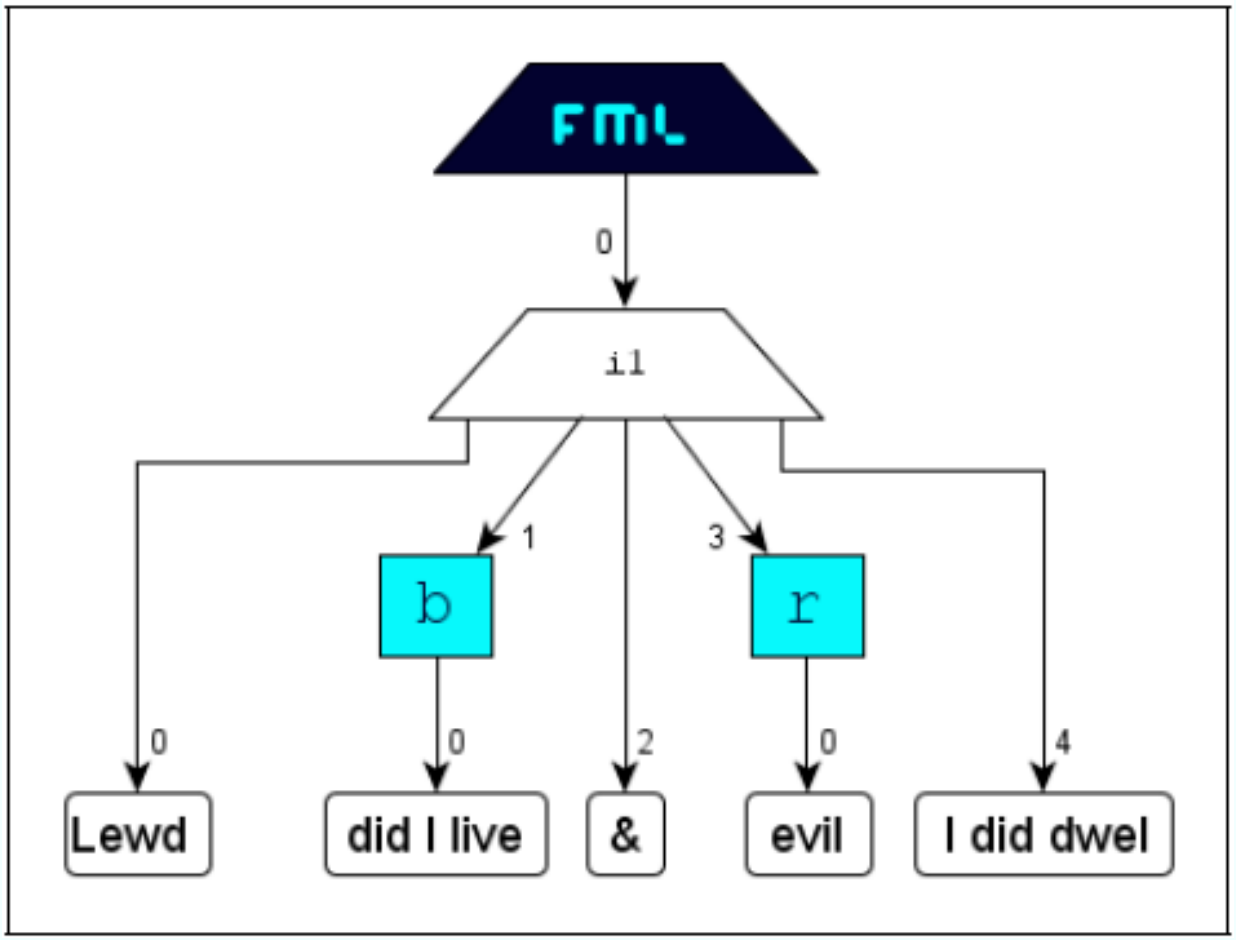
CONCEPTS: independence



CONCEPTS: independence

```
Lewd <i1|b>did I live<i1|/b> & <i1|r>evil<i1|/r> I did dwel
```

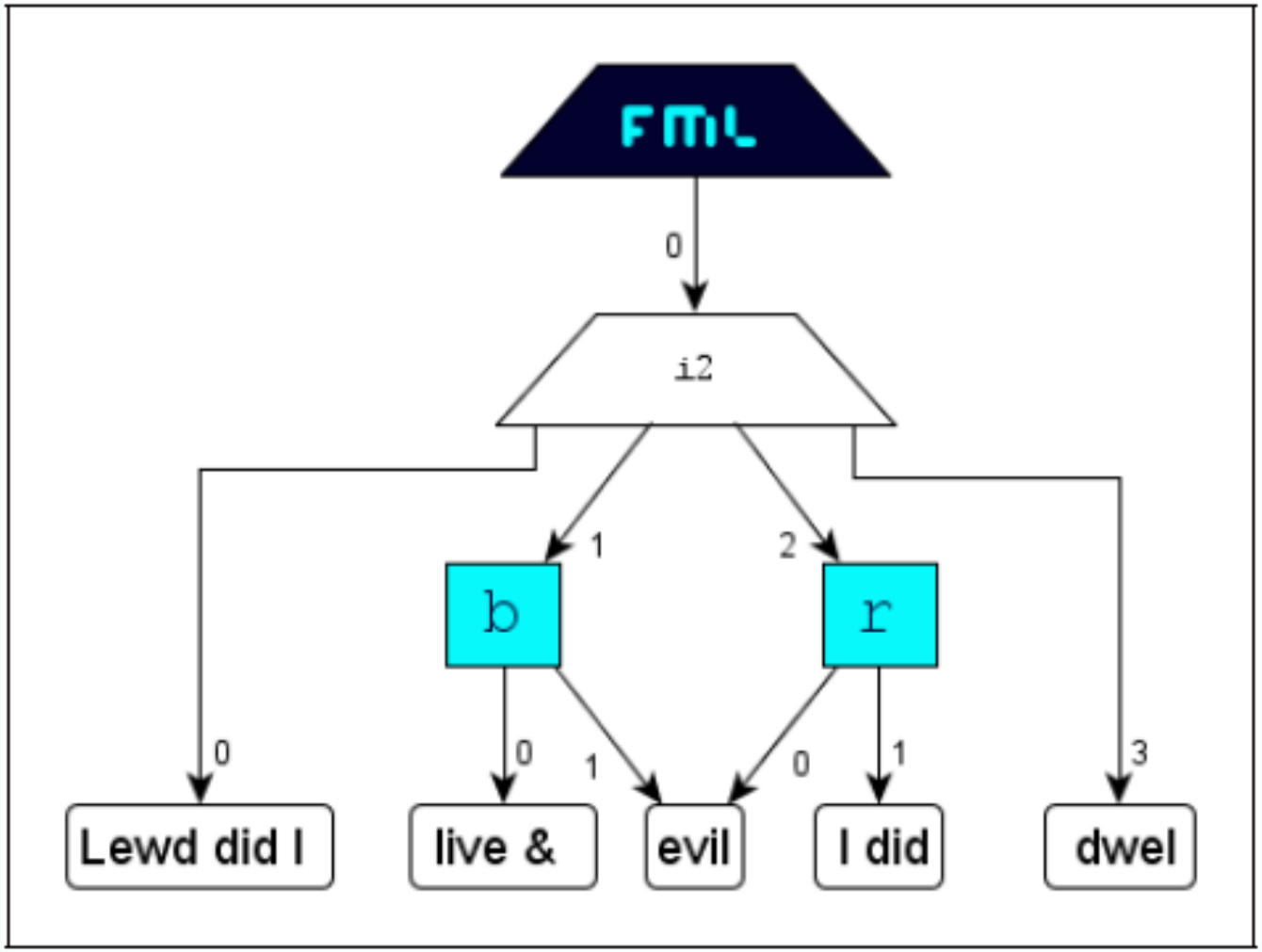
CONCEPTS: independence



CONCEPTS: independence

```
Lewd did I <i2|b>live & <i2|r>evil<i2|/b> I did<i2|/r> dwel
```

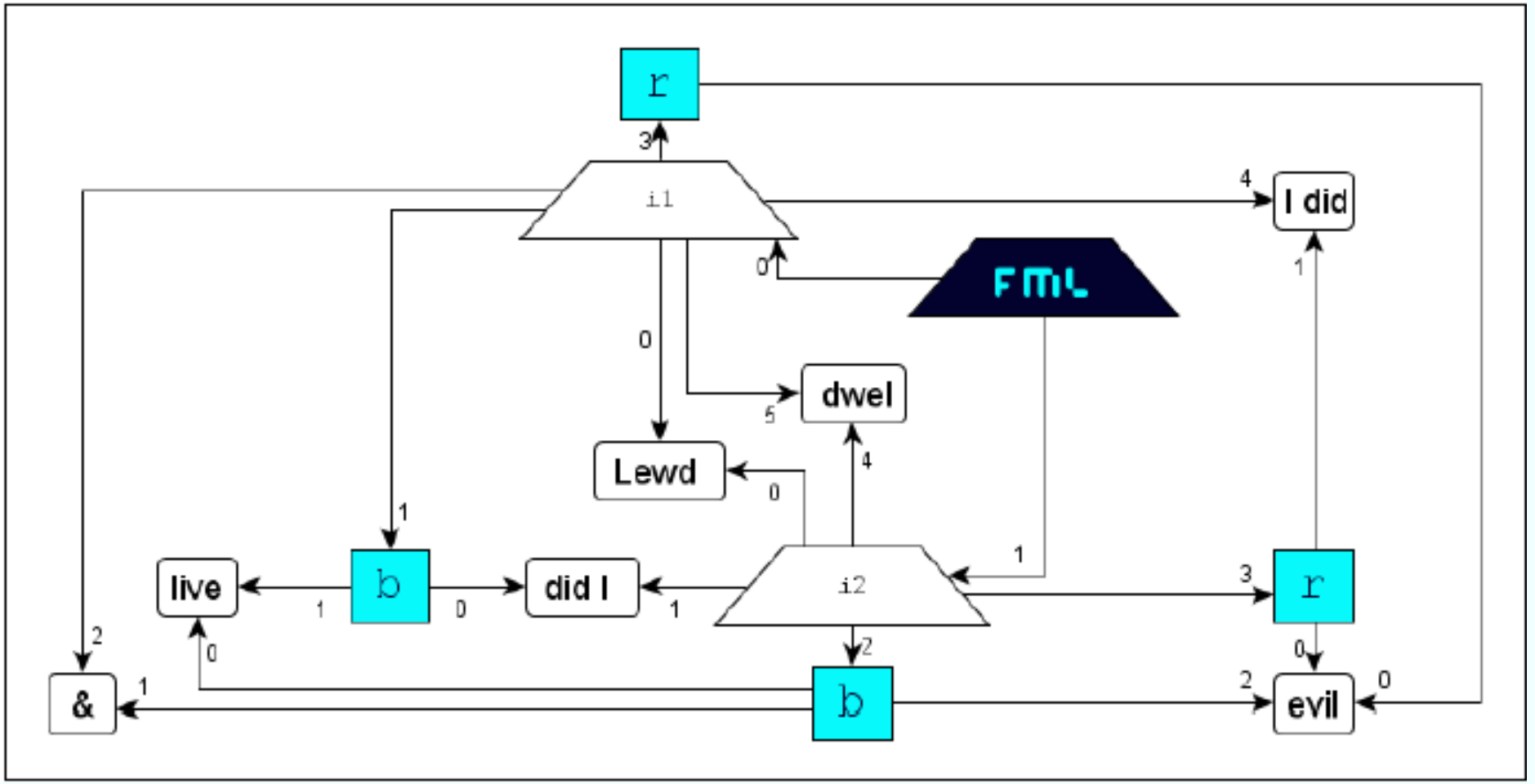
CONCEPTS: independence



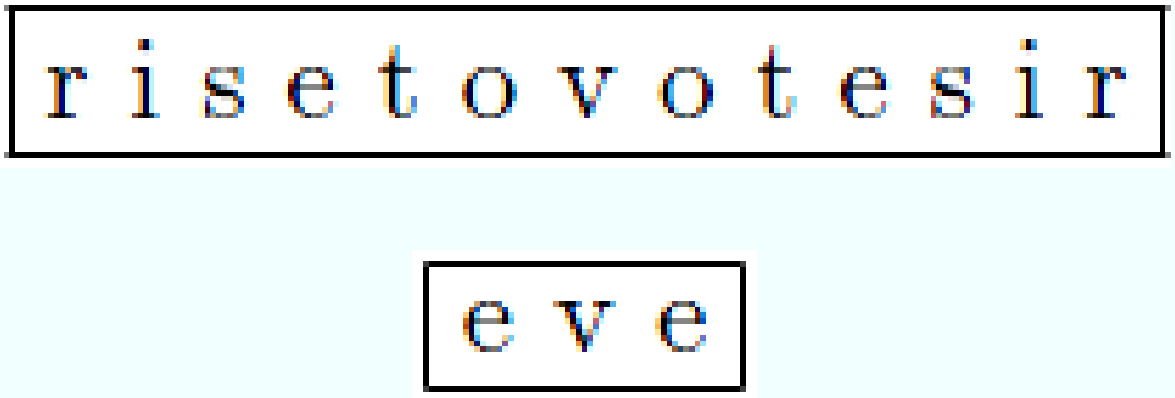
CONCEPTS: independence

```
Lewd <i1|b>did I <i2|b>live<i1|/b> & <i2|r><i1|r>evil
<i1|/r><i2|/b> I did<i2|/r> dwel
```

CONCEPTS: independence



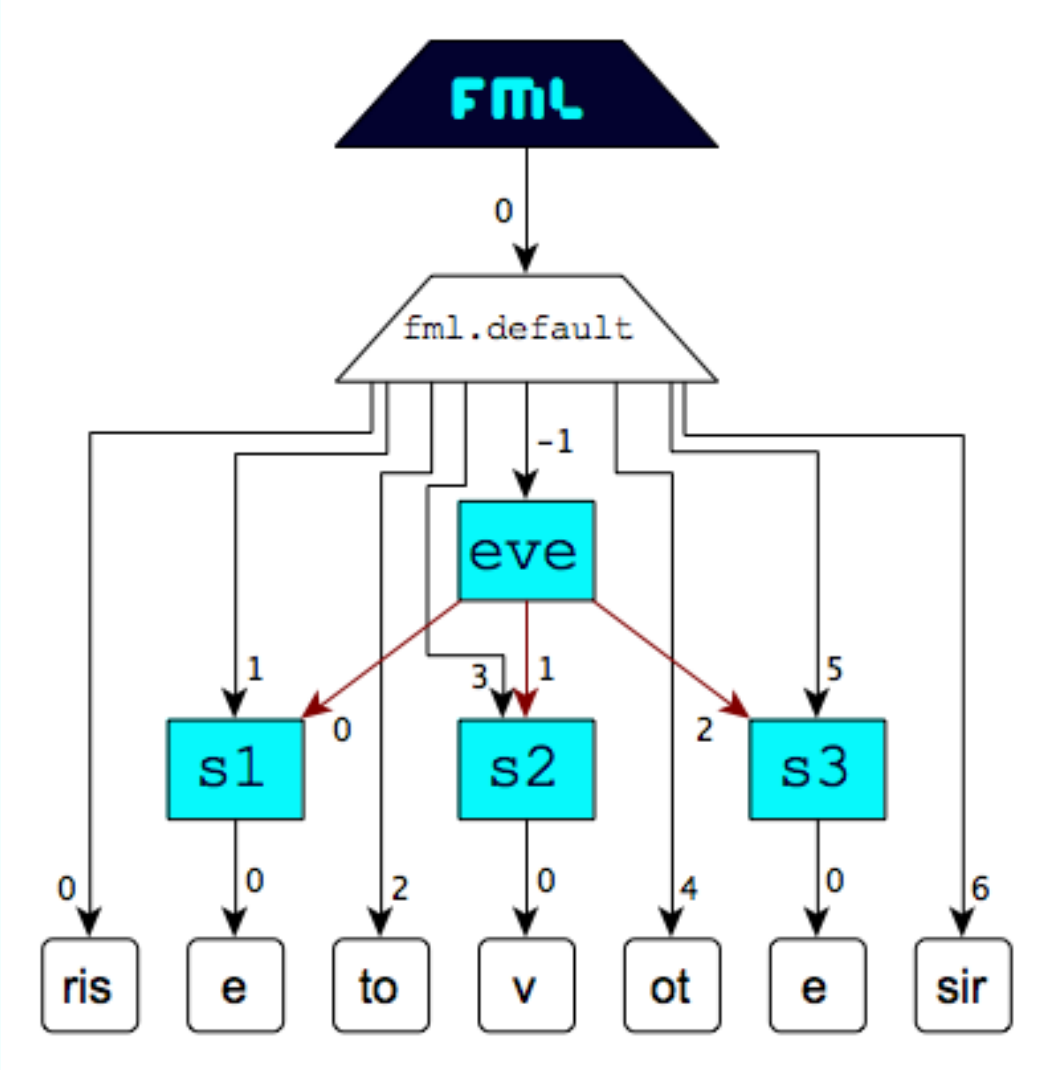
CONCEPTS: segmentation



CONCEPTS: segmentation

```
ris
<s1 fml.segment.id="eve" fml.segment.pos="0">
    e
</s1>
to
<s2 fml.segment.id="eve" fml.segment.pos="1">
    v
</s2>
ot
<s3 fml.segment.id="eve" fml.segment.pos="2">
    e
</s3>
sir
```

CONCEPTS: segmentation



CONCEPTS: segmentation

Rank ^	Start No	Bib	Country	Name	Turns Score	Air Score	Time Score	Score
1	19	4		BILODEAU Alexandre	14.1	5.44	7.21	26.75
2	17	1		BEGG-SMITH Dale	14.2	5.43	6.95	26.58
3	18	5		WILSON Bryon	13.8	5.46	6.82	26.08

CONCEPTS: segmentation

Canada	BILODEAU Alexandre	26.75
Australia	BEGG-SMITH Dale	26.58
United States	WILSON Bryon	26.08

$$f : Z \times S \to I$$

$$f : (1,1) \mapsto \texttt{Canada},$$
$$f : (1,2) \mapsto \texttt{BILODEAU Alexandre},$$
$$f : (1,3) \mapsto 26.75,$$
$$f : (2,1) \mapsto \texttt{Australia},$$
$$f : (2,2) \mapsto \texttt{BEGG-SMITH Dale},$$
$$f : (2,3) \mapsto 26.58,$$
$$f : (3,1) \mapsto \texttt{United States},$$
$$f : (3,2) \mapsto \texttt{WILSON Bryon},$$
$$f : (3,3) \mapsto 26.08$$

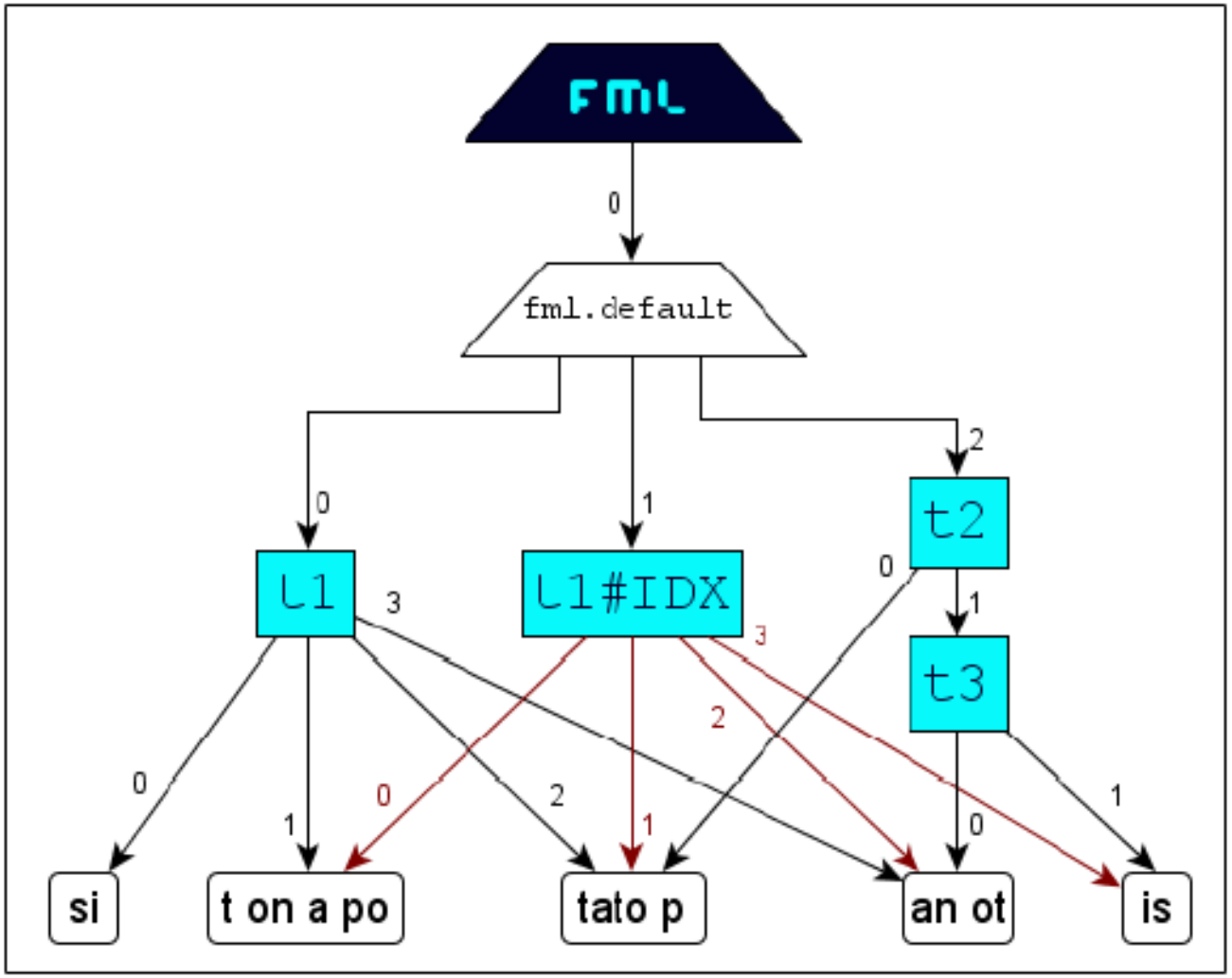
CONCEPTS: fragmentation

```
si<t1 #IDX>t on a po<t2>tato p<t3>an ot</t1>is
```

CONCEPTS: fragmentation

```
01 <t1>
02 si<t1#IDX>t on a po<t2>tato p<t3>an ot</t1>is
03 </t1#IDX /t2 /t3>
```

CONCEPTS: fragmentation

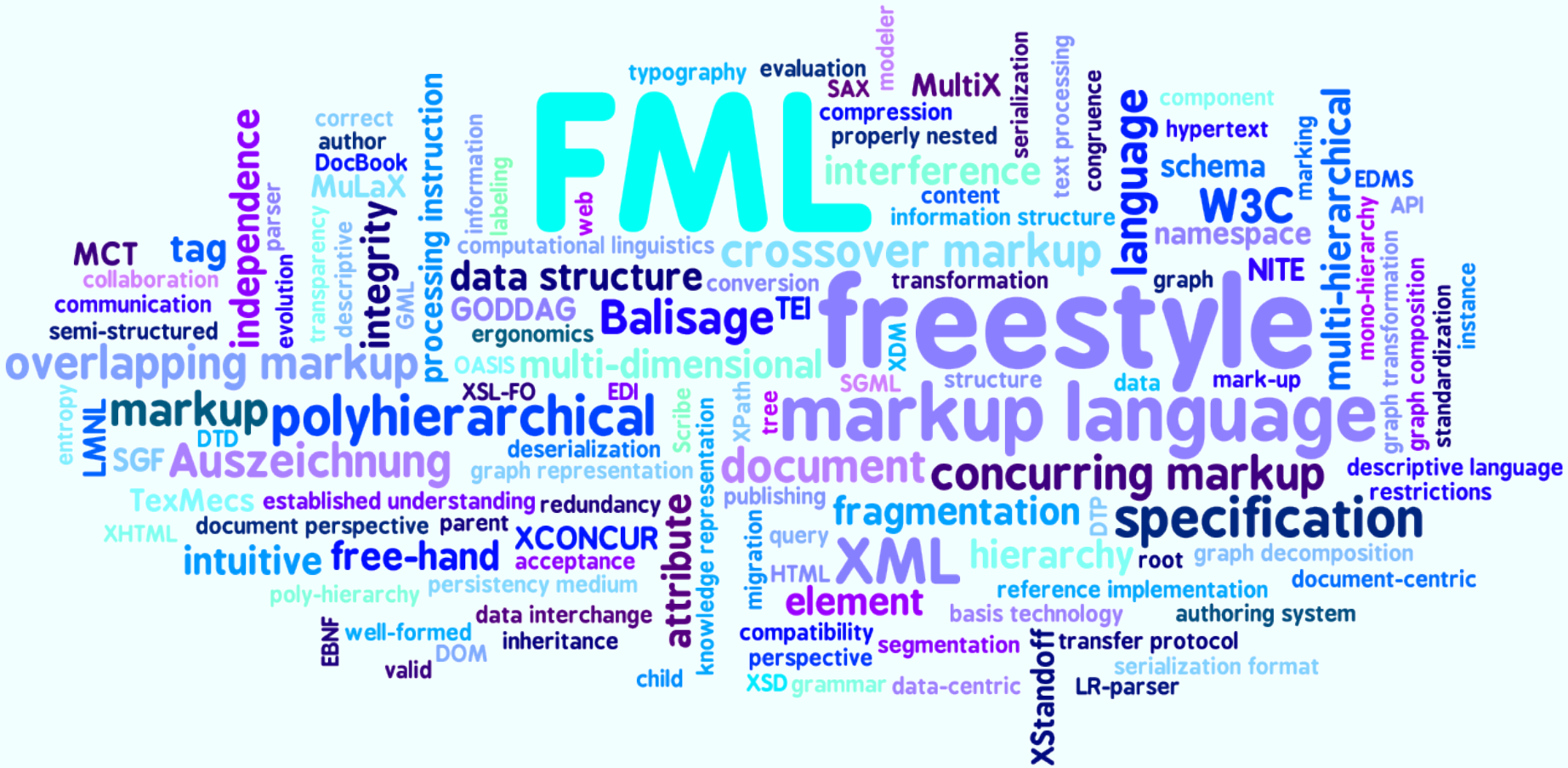


STATUS

- ✓ reference analysis
- ✓ Freestyle Document
- ✓ Freestyle Graph
- ✓ transformation guidelines
- ✓ XML representation
- ✓ API

FUTURE WORK

- verification
- reference implementation
- Freestyle Editor
- www.freestyle-markup.org maintenance
- 2ndary technologies...


FML
freestyle
markup language
crossover markup
overlapping markup
polyhierarchical
Balisage
Auszeichnung
concurring markup
specification
fragmentation
document
XML
hierarchy
element
attribute
free-hand
intuitive
XCONCUR
data structure
multi-dimensional
multi-hierarchical
language
interference
independence
integrity
processing instruction
markup
tag
MCT
W3C
namespace
schema
hypertext
component
MultiX
compression
properly nested
typography
evaluation
SAX
modeler
serialization
text processing
congruence
content
information structure
transformation
conversion
GODDAG
ergonomics
TEI
NITE
graph
data
mark-up
structure
SGML
XDM
XSL-FO
EDI
OASIS
deserialization
graph representation
TexMecs
established understanding
redundancy
document perspective
parent
XHTML
acceptance
persistency medium
poly-hierarchy
data interchange
inheritance
well-formed
DOM
valid
EBNF
child
knowledge representation
Scribe
XPath
tree
publishing
migration
query
HTML
DTP
root
graph decomposition
reference implementation
basis technology
document-centric
authoring system
compatibility
perspective
segmentation
transfer protocol
serialization format
XSD
grammar
data-centric
LR-parser
XStandoff
descriptive language
restrictions
standardization
instance
graph composition
graph transformation
mono-hierarchy
EDMS
API
marking
collaboration
communication
semi-structured
evolution
parser
transparency
descriptive
GML
information
labeling
web
correct
author
DocBook
MuLaX
entropy
LMNL
SGF
DTD